



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2016

Exploratory Study on Managing and Searching for Documents in Software Engineering Environments

Ben Charrada, Eya ; Mussato, Stefan

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: <https://doi.org/10.5167/uzh-136249>
Conference or Workshop Item

Originally published at:

Ben Charrada, Eya; Mussato, Stefan (2016). Exploratory Study on Managing and Searching for Documents in Software Engineering Environments. In: ACM Symposium on Document Engineering, Vienna, 13 September 2016 - 16 September 2016, s.n..

An Exploratory Study on Managing and Searching for Documents in Software Engineering Environments

Eya Ben Charrada
Institute for Informatics
University of Zurich
Switzerland
charrada@ifi.uzh.ch

Stefan Mussato^{*}
Kaba AG
Rümlang
Switzerland
stefan.mussato@dormakaba.com

ABSTRACT

A large number of documents are usually produced in the software industry. In this work, we conduct a qualitative study to explore the main practices and challenges related to managing these documents. The results of this study are based on interviews with 13 practitioners from nine companies. The main findings of the study are: (1) much data is stored in e-mails and in meeting protocols, (2) practitioners like wikis, (3) when searching for documents, practitioners would rather browse the structure than use the search function and (4) searching for documents is still a challenge due to the low effectiveness of search functions and the scattering of documents over several locations and tools.

Keywords

Document management; software engineering; qualitative study; industrial study

1. INTRODUCTION

In software development projects, a large number of documents, technical and non-technical, are usually created and used by practitioners with various backgrounds and positions (managers, developers, sale, etc.). For these practitioners, managing the documents is a daily task that can be more or less challenging. In this work, we aim at exploring the practices and challenges related to document management in the software industry. Unlike previous studies and surveys that have been conducted in this field (e.g. [5, 3, 7]), this work has a special focus on how documents are searched for. Furthermore, previous studies usually focused on the use of technical software documentation by software engineers, while this work considers all documents that are produced and used by practitioners holding various technical and non-technical positions.

The results of this work are based on semi-structured interviews with 13 participants working in software development environments. The participants are from nine companies, all based in Switzerland.

^{*}This work was conducted during the author's studies at the University of Zurich

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DocEng'16, September 12–16, 2016, Vienna, Austria.

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4438-8/16/09...\$15.00

DOI: <http://dx.doi.org/10.1145/2960811.2967149>

The remainder of the paper is organized as follows. In Section 2, we present the research question, design of the study, and threats to validity. We report the main results of the study in Section 3 and discuss them in Section 4. In section 5, we present related studies and surveys and we conclude in Section 6.

2. RESEARCH METHODOLOGY

To understand the practices of document management in software development environments, we conducted an exploratory study with nine companies. Since we aim at investigating new and diverse data, we chose a qualitative research approach. The study was conducted via semi-structured interviews. In the remainder of this section, we present the research question and the study design.

2.1 Research question and topics of interest

The study aims at answering the following question:

How do practitioners manage and search for documents in software environments and what challenges they face?

Through this question, we would like to investigate the following aspects. Primarily, we would like to explore how practitioners, from different positions, search for documents. By document we mean any data that is stored electronically on storages provided by the company or on the practitioners' own storages. In addition to searching for documents, we are also interested in how practitioners manage (e.g. create, store and structure) documents and how this impacts their search strategies. Furthermore, we would like to explore to what extent is the ease of search impacted by the companies' processes and guidelines for document management. Finally, we are interested in identifying the main challenges that practitioners face when searching for documents and what limitations they see in current document management tools. Practitioners from the software engineering field are likely to be fairly familiar with software tools and technologies and it is therefore interesting to find out the main problems they face.

2.2 Initial preparation

After reviewing the literature related to document management in software engineering, we defined an initial set of topics of interest. We then prepared questions that we used to conduct five pilot interviews with practitioners. The pilot interviews covered various aspects about documentation such as the type of documents that practitioners create and why, usages of documents in practice and the challenges that practitioners face with current document management practices. The output of the pilot interviews was then used for refining the topics of interest and redefining the questions to be used during the interviews. The output of the pilot interviews is not included in the results reported in this paper.

2.3 Selection of participants and data collection

Thirteen interviewees from nine companies all based in Switzerland were contacted and interviewed. The sector in which the companies are working are presented in Table 1. All interviewees are from the personal network of the second author. Since the number of participants is not too large and as we would like to get a representative picture about the practices of document management, the sampling was done towards covering various positions within software projects. The positions held by the interviewees as well as their experience are presented in Table I. All interviews were organized, conducted and transcribed by the second author, a product manager with 23 years experience in industry who was also enrolled as student at the University of Zurich during the interviewing period. The interviews took place from April to June 2015. All interviews were conducted face to face and were audio recorded after getting the consent of the interviewees. The interview language was German, which is the mother tongue of the participants.

2.4 Data transcription and analysis

The interviewer translated the interviews to English while transcribing them. The transcription was not done word to word but was a crisp reformulation of what the participants said. The analysis was mainly performed by the first author based on the transcribed data. In case of doubt during the analysis or when looking for quotes, the author used the audio recordings to check the original data. Although the collected data covered aspects related to search for information in general (e.g. in the internet), this data was excluded during the analysis so we only focused on data related to searching within documents stored electronically in the company storages or in the practitioner's own storages.

2.5 Threats to validity

Qualitative studies are subject to several threats. Below we explain the design decisions that we made to limit potential ones.

Construct Validity issues appear if the answers obtained from the participants do not accurately reflect the real practice. In order to limit this threat, we informed the participants about the anonymity of the study and we avoided judgements during the interviews. Misunderstandings between the interviewer and interviewees constitute another threat to construct validity. In order to minimize this, the interviewer explicitly explained the goal of the study in advance. Additionally, all studies have been conducted in German which is the mother tongue of both the interviewer and the interviewees. Since the person analysing the data is different from the person conducting and transcribing the interviews, there is a risk of misinterpreting the transcribed data. In order to limit this risk, the analyst listened to the audio files whenever there was doubt about the meaning or context of the transcribed data. Additionally, the interviewer was involved in checking the results that the analyst produced.

Since all participants are from the network of the interviewer, no one declined. This limits the threat of getting only participants that have a special interest in the topic, which would present an *internal validity* threat. However, getting all the participants from the same network could result in a group of participants that have similar backgrounds. To reduce this threat, we selected participants with heterogeneous positions (e.g. technical, managerial) from companies working in various sectors (e.g. manufacturing, consultancy, etc.).

Since the study is qualitative, generalizing the results beyond the settings of this study is rather difficult (*external validity*). Nev-

ertheless, since the selected participants had diverse positions and settings that would cover the main roles in software development projects, findings that apply to the different participants of the study are likely to apply to other practitioners in software development environments.

3. RESULTS

Searching for documents and data is a daily task for software practitioners that takes much of their time. Although practitioners use and manage documents in diverse ways, some similar practices and challenges were reported by many interviewees. In the reminder of this section, we report the main findings we observed in this study.

E-mails and protocols store data

Although not primarily meant for data storage, engineers use e-mails for storing data such as agreements information, reports, documents and links. Example of data looked for within e-mails are:

"...what was done, what was already addressed and how it was addressed, what was decided." – P2

Protocols or slides of meetings are also among the frequently searched documents among managers. When asked about what kind of information he looks for, a participant answered:

"...a lot of protocols or e-mails, I know that people had sent me that and that, and I need to look again how that was..." – P1

Meeting protocols are mostly office documents, but could also be found as plain text in e-mails. To facilitate information finding within e-mails several practitioners (seven in our case) create their own, more or less detailed, structure, e.g. based on projects. However, due to the growing number of e-mails they receive, two participants mentioned giving up on e-mail categorization and switching to solely relying on the search function of the e-mail program.

"...I had a huge folder system with defined themes where I store and find e-mail in a relatively good way, but it is so that I have since long lost the fight... I receive so many e-mails..." – P1

Finding information within e-mails is usually regarded by practitioners as easy, unlike the search within other types of storages, which could be very challenging (this point will be discussed further later). One factor influencing the ease of search in e-mails is the good capabilities of the search function in e-mail clients, which participants were generally happy with. Additionally, participants usually remember data and meta-data about the e-mail such as the sender name, keywords from the e-mail subject and body or the project number. Only three participants mentioned deleting e-mails on a regular basis. One of them does it as soon as the task is done, while the two others keep the e-mail in a folder structure for some time before deleting them.

Practitioners like wikis

Wikis were mentioned in eleven interviews. Among the eleven interviewees, only one, with a managerial position, reported never having used a wiki. The other participants had different levels of experience with wikis. Wikis seem to be very liked by practitioners with technical tasks

"...[the wiki is used] for the support, for example to get information about recurrent problems and failures, and the development of course, and also the programming." – P7

Table 1: Characterization of the companies and participants

Company ID	Size ^a	Sector	Participant ID	Experience ^b	Position
C1	L	Manufacturing industry	P1	26	Product manager
C1	L	Manufacturing industry	P2	30	Program Lead
C1	L	Manufacturing industry	P3	25	Technical sales manager
C1	L	Manufacturing industry	P4	25	Business owner
C2	L	Consulting and Engineering	P5	19	Consultant
C2	L	Consulting and Engineering	P6	>10	Consultant
C3	L	Banking	P7	16	IT security manager
C4	S	IT security	P8	18	Security manager
C5	L	Consulting	P9	13	Consultant/ Requirements Engineer
C6	S	IT software services	P10	20	Company owner (One person business)
C7	L	Manufacturing Industry	P11	23	Department head
C8	M	Interaction Design	P12	19	Interaction designer
C9	M	Software solutions	P13	30	System Engineer

^a Company size: (S)mall, (M)edium, (L)arge^b Experience in years

“...it is thin and quick...it has surely not a great look and feel, it contains sequences of information and as a technical person I find this good” – P13

But also practitioners with less technical positions seem to prefer wikis to other content and document management systems, and this is mainly for its better structure and better search function:

“...the search function of the wiki simply provides better results...” – P9

Other cited advantages of wikis are that they allow a simple and fast way of changing the content in the same media, and that the relevant pages in wikis can be marked in the browser bookmarks which facilitates finding them later on. Only one participant mentioned not liking the wiki, and this is because of its structure:

“I don’t like wikis too much; there’s a large amount of data and a strange structure ... I often have the impression of document cavers” – P12

Browsing first, then using the search function

When looking for a specific document, practitioners would rather browse the folder structure than use the search function. Therefore, to facilitate the search for data, many participants mentioned defining a good structure on their own drives and some on their e-mail clients:

“I have a clever storage [using a catalogue by topics] which makes searching obsolete” – P2

“If I do not find a document or information at the first location, I would retry in another folder. If still there is no match, I may attempt to use the search on all locations ...” – P4

When looking within a collaborative structure, finding documents by browsing becomes more challenging. Practitioners would then behave differently depending on whether they have an idea or not about the location of the document. In the former case, practitioners use browsing to find the document. In the latter case, we identified two types of behaviour. The majority of users find the search function to be unreliable and thus still try to guess where the document is and use browsing. For these users the search function is only used as a last resort if no results could be found by browsing.

“I had higher expectations with the [shared storage] tool, the search is (thinking), well, the logic is not so good, when not knowing exactly what one is looking for it is not reliable. So I try to remember in which folder and in which structure is the file.” – P1

A few users, however, wouldn’t mind using the search function and might even find it faster than browsing in that case.

“I don’t use the catalogue [browsing] in the shared storage tool because it does not have a logic structure...I can only use the search...I am faster then” – P2

“[When looking within shared structures, I use] mostly structure browsing too, but I also use the search function more often” – P4

As browsing is the preferred way for searching, the structure of the shared data and the company processes about document management influence the efficiency and ease of the search. Regarding the search for documents within shared drives, a participant mentioned:

“Because the structure is pre-defined and each project has its structure, it is relatively simple to find...” – P12

In practice, however, only about half the participants mentioned having a more or less strict processes for document storage, while the others only have loose or no regulations.

Searching for documents: the challenges

Although searching for documents is a very common and recurrent task, it is not always easy. In fact, several interviewees mentioned having experienced problems and frustration when looking for a document:

“...I know that it exists, I have opened it once, but after that I didn’t find it anymore.” – P9

“That’s very annoying, then depending on how important the document is, I might ask someone... I might also reproduce it again...” – P12

Several factors have been cited as negatively impacting the search for documents in practice. In some cases, companies would use more than one tool for document storage, which results in the data being scattered over multiple locations. Consequently, when not sure about the location of a document, practitioners need to try

the search on each location separately. Some practitioners might even have their own storage on top of that.

“...I use Google Docs from one project and then Dropbox and now OneNote, then my computer and the company computer...I need to shortly think about which project is it, where do they work and why...” – P12

Available search tools are also considered as very limited by most participants. This is because they deliver too many matches, or in some cases, no match:

“I also experienced situations where a search function did not show the expected match even though the keyword entered was correct. Of course this disappointed me...” – P4

One participant mentioned that the search logic which is different among different tools would require more effort for performing the search. Additional factors that were mentioned to negatively influence the data search are human-errors such as the document names not matching content, or using a different spelling in the document name.

4. DISCUSSION

Morkala and Maurer [6] identified scattered documentation and limited search functions as communication wastes in globally distributed agile software projects. Our results support this finding since scattering of documents over several tools has also been reported to hinder document search. However, in our study this problem was not linked to globally distributed teams. In fact, such a problem happens when the company (or even the practitioner) uses more than one storage tool.

When browsing the structure, practitioners need to rely much on memory and/or guessing to find the location of a document. Therefore, the search would be easier if the practitioner has prior knowledge about the document location. This finding is in accordance with the results of the study of de Graaf et al. [1], which identified that prior knowledge helps in searching software documentation efficiently and effectively.

Despite the advances made in document search, current search tools do not seem to meet the expectations of practitioners. In fact, many practitioners see the offered search functionalities as limited, unreliable, and lead to all kinds of results. Consequently, many practitioners use structure browsing as the main search strategy even when the location of the document is not known for them. Since the search space is usually huge, practitioners need to guess where could the file be located, which is not always effective. There is therefore a clear need for better document management and search tools that are aligned with the users' behaviour and way of thinking.

5. RELATED WORK

Studies about the management and use of technical documentation in software projects are numerous. While some of these studies cover very specific tasks such as architectural design [8] or software maintenance [2], others are more generic and cover the use of various types of documentation [3]. Due to space limit, we restrict our focus to recent industrial studies covering aspects that overlap with the scope of this work. The work of Garousi et al. [5] explored how technical documentation is used and what factors impact its usefulness in practice. Plösch et al. [7] assessed the most important

quality attribute of software documentation via a survey. Related to search of information, Freund [4] explored what contextual factors impact the information seeking behaviour of software engineers. The focus of our work, which is storage and search for documents, is different from the previous studies. Furthermore, the scope of this work is also larger since it covers not only software documentation but also other type of documents produced in software environments.

6. CONCLUSION

In this study, we explored practices and challenges related to managing documents in software development environments. Our findings include strategies that practitioners use to structure and search for documents. We also report challenges that practitioners face when searching for documents. For future work, we plan to use the findings of this study to conduct a survey that will provide quantitative data on the topic.

Acknowledgments

We would like to thank all the participants for accepting to participate, and for their time.

7. REFERENCES

- [1] De Graaf, Klaas Andries, Peng Liang, Antony Tang, and Hans Van Vliet. The impact of prior knowledge on searching in software documentation. In *Proceedings of the 2014 ACM symposium on Document engineering*, pages 189–198. ACM, 2014.
- [2] de Souza, Sergio Cozzetti B., Nicolas Anquetil, and Kathia M. de Oliveira. A study of the documentation essential to software maintenance. In *Proceedings of the 23rd annual international conference on Design of communication: documenting and designing for pervasive information*, pages 68–75. ACM, 2005.
- [3] Forward, Andrew, and Timothy C. Lethbridge. The relevance of software documentation, tools and technologies: A survey. In *Proceedings of the 2002 ACM symposium on Document engineering*, DocEng '02, pages 26–33. ACM, 2002.
- [4] Freund, Luanne. Contextualizing the information-seeking behavior of software engineers. *Journal of the Association for Information Science and Technology*, 66(8):1594–1605, 2015.
- [5] Garousi, Golara, Vahid Garousi-Yusifoglu, Guenther Ruhe, Junji Zhi, Mahmoud Moussavi, and Brian Smith. Usage and usefulness of technical software documentation: An industrial case study. *Information & Software Technology*, 57:664–682, 2015.
- [6] Korkala, Mikko, and Frank Maurer. Waste identification as the means for improving communication in globally distributed agile software development. *Journal of Systems and Software*, 95:122–140, 2014.
- [7] Plösch, Reinhold, Andreas Dautovic, and Matthias Saft. The value of software documentation quality. In *14th International Conference on Quality Software*, pages 333–342. IEEE, 2014.
- [8] Rost, Dominik, Matthias Naab, Crescencio Lima, and Christina von Flach Garcia Chavez. Software architecture documentation for developers: a survey. In *European Conference on Software Architecture*, pages 72–88. Springer, 2013.